# Corrigé TD 4
Langages Formels

Luc Lapointe
luc.lapointe@ens-paris-saclay.fr
home.lmf.cnrs.fr/LucLapointe/

## Exercice 1

Soit $\mathcal{L}$ un langage régulier. Montrer que $\{u\tilde{v} \mid uv \in \mathcal{L}, u \neq v\}$ est algébrique, où $\tilde{v}$ est le mot miroir de $v$.

**Idée de solution**

Si $\mathcal{L}$ est régulier, $\{u\tilde{v} \mid uv \in \mathcal{L}\}$ est régulier. Pour le montrer, considérer $\mathcal{A}$ l'automate qui reconnaît $\mathcal{L}$, et $\mathcal{A}_{q_I \to q}$ et $\mathcal{A}_{q \to q_F}$ les automates qui sont identiques à $\mathcal{A}$ mais respectivement ont comme unique état final $q$ et état initial $q$, puis appliquer le miroir aux $\mathcal{A}_{q \to q_F}$.

Pour éliminer les cas où $u = v$, empiler les lettres lues dans les $\mathcal{A}_{q_I \to q}$, et les dépiler dans les $\mathcal{A}_{q \to q_F}$.

## Exercice 2 – Parenthèses retournées

Soit $D_n$ le langage de Dyck à n paires de parenthèses. Les langages suivants sont-ils algébriques ?

- $\{u\tilde{v} : uv \in D_1\}$
- $\{u\tilde{v} : uv \in D_2\}$

**Idée de solution**

- Commençons par remarquer que $D_1$ est généré par

$$S \to SS \mid (S) \mid \varepsilon$$

et $\widetilde{D_1}$ par

$$S \to SS \mid )S( \mid \varepsilon$$

Lorsqu'on factorise un mot $w \in D_1$ en $w = uv$, certaines parenthèses de $u$ ne sont pas fermées, et certaines de $v$ pas ouvertes. Si $u$ a $n$ parenthèses non fermées, $u$ est de la forme

$$u = u_0(u_1(...u_{n-1}(u_n$$

avec $u_0...u_n \in D_1$, et $v$ a une forme

$$v = v_0)v_1)...v_{n-1})v_n$$

donc notamment $\tilde{v}$ est de la forme

$$\tilde{v} = \tilde{v}_n)\tilde{v}_{n-1})...\tilde{v}_1)\tilde{v}_0$$

Nous pouvons en déduire que le langage recherché est généré par

$$S \to X(S)Y \mid XY$$
$$X \to XX \mid (X) \mid \varepsilon$$
$$Y \to YY \mid )S( \mid \varepsilon$$

où $X$ génère des mots de $D_1$ et $Y$ de $\widetilde{D_1}$.

- Considérer la famille de mots inclus dans le langage recherché

$$(^N[^N)^N]^N$$

## Exercice 3 – Distance de Hamming

La distance de Hamming entre deux mots de même longueur est le nombre de positions où les mots diffèrent (ex: *abac* et *abba* ont une distance de 2). Pour tout langage régulier $\mathcal{L}$, montrez que le langage des mots $v$ à distance au plus $|v|/2$ d'un mot de longueur $|v|$ de $\mathcal{L}$ est algébrique. Montrez qu'il n'est pas toujours régulier.

**Idée de solution**

Soit $\mathcal{M}$ le langage recherché. Considérer un automate $A$ qui reconnaît $L$, et construire à partir de $A$ un automate à pile qui reconnaît $\mathcal{M}$. Cet automate compte les positions sur lesquelles les lettres sont identiques, et celles sur lesquelles non.
Pour chaque transition $q \xrightarrow{a} q'$ dans $A$, ajouter une transition à $\mathcal{M}$ qui rajoute un symbole $+$ en lisant un $a$ sur une pile vide ou une tête de pile $+$, ou supprime une tête de pile $-$; et qui rajoute un $-$ en lisant une lettre différente de $a$ sur une pile vide ou une tête de pile $-$, ou supprime une tête de pile $+$. Accepter à la fin de la lecture du mot seulement s'il n'y a aucun $-$ sur la pile.

Le langage décrit n'est pas toujours régulier. Considérer $a^*$ sur l'alphabet $\Sigma = \{a, b\}$. Le langage avec plus de $a$ que de $b$ n'est pas régulier.

## Exercice 4 – ($\star$) Theorème de Parikh

Soit $\Sigma = a_1, ..., a_n$. L'image d'un mot $w$ est le tuple $|w|_{a_1}, ..., |w|_{a_n} \in \mathbb{N}^n$ . L'image d'un langage est l'ensemble des images de ses mots. Montrer que pour tout langage algébrique, il existe un langage régulier avec la même image.

**Preuve**

La page Wikipedia en français du théorème recense différentes preuves. Faites votre choix ! Notamment, vous pouvez essayer de lire la preuve issue du papier original.

# Contrôle continu

*N'hésitez pas à réfléchir à cet exercice à plusieurs, du moment que vous rédigez uniquement des choses que vous comprenez, et avec vos propres mots.*

Soit $\mathcal{L}$ un langage algébrique tel que pour toute paire de mots de $\mathcal{L}$, l'un est préfixe de l'autre. Le but de cet exercice est de montrer que $\mathcal{L}$ est régulier.

1. Montrer que $\mathcal{L}$ ne peut pas contenir plusieurs mots différents de même taille.
2. Grâce au lemme de pompage, montrer que $\mathcal{L}$ est inclus dans un langage de la forme $\{xy^n \mid n \in \mathbb{N}\}$ où $x$ et $y$ sont des mots.
3. Montrer que $\mathcal{L}$ est l'intersection de deux langages réguliers, inspirés des deux questions précédentes.